

Research Article

Computation of Invariant Measures and Stationary Expectations for Markov Chains with Block-Band Transition Matrix

Hendrik Baumann ¹ and Thomas Hanschke ²

¹Clausthal University of Technology, Institute of Mathematics, Erzstr. 1, 38678 Clausthal-Zellerfeld, Germany

²Simulation Science Center Clausthal-Göttingen, Arnold-Sommerfeld-Str. 6, 38678 Clausthal-Zellerfeld, Germany

Correspondence should be addressed to Hendrik Baumann; hendrik.baumann@tu-clausthal.de

Received 14 January 2020; Revised 13 March 2020; Accepted 23 March 2020; Published 8 July 2020

Academic Editor: Mirosław Lachowicz

Copyright © 2020 Hendrik Baumann and Thomas Hanschke. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper deals with the computation of invariant measures and stationary expectations for discrete-time Markov chains governed by a block-structured one-step transition probability matrix. The method generalizes in some respect Neuts' matrix-geometric approach to vector-state Markov chains. The method reveals a strong relationship between Markov chains and matrix continued fractions which can provide valuable information for mastering the growing complexity of real-world applications of large-scale grid systems and multidimensional level-dependent Markov models. The results obtained are extended to continuous-time Markov chains.

1. Introduction

This paper is concerned with the computation of invariant measures of discrete-time Markov chains $X = (X_k)_{k \geq 0}$ having a block-partitioned one-step transition matrix P of the form

$$P = \begin{pmatrix} P_{00} & P_{01} & 0 & 0 & \cdots & \cdots & \cdots & \cdots \\ P_{10} & P_{11} & P_{12} & 0 & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & & & & \\ P_{n0} & P_{n1} & \cdots & P_{nn} & P_{n,n+1} & 0 & 0 & \cdots \\ 0 & P_{n+1,1} & \cdots & P_{n+1,n} & P_{n+1,n+1} & P_{n+1,n+2} & 0 & \cdots \\ \vdots & \ddots & & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \quad (1)$$

where all elements are nonnegative matrices of suitable dimension, that is, $P_{in} \in \mathbb{R}^{d_i \times d_n}$ for $d_0, d_1, \dots \in \mathbb{N} = \{1, 2, 3, \dots\}$. The states of the Markov chain are denoted by (i, j) , $i \geq 0$ and $1 \leq j \leq d_i$, and are ordered in lexicographic order,

that is, $(0, 1), \dots, (0, d_0), (1, 1), \dots, (1, d_1), (2, 1), \dots$. Following Neuts [1], the set of states $\{(i, 1), \dots, (i, d_i)\}$, $i \geq 0$, will be called the *level* i . We refer to the first component of X_k (the level) as $X_k^{(1)}$, and by $X_k^{(2)}$, we denote the second component of X_k . Throughout this paper, it is assumed that $n \geq 1$ and that the Markov chain is irreducible and recurrent. Note that irreducibility implies $P_{i,i+1} \neq 0$ for all $i \in \{0, 1, 2, \dots\}$. From the general theory of Markov chains, we then know that the system

$$\mathbf{x}P = \mathbf{x}, \mathbf{x} \geq 0, \quad (2)$$

has a unique solution (up to a constant factor) which is strictly positive (see Karlin and Taylor ([2], chapter 11)). For the solution vector \mathbf{x} , we use the notation $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots)$ and $\mathbf{x}_i = (x_{(i,1)}, x_{(i,2)}, \dots, x_{(i,d_i)})$, $i \geq 0$. Every constant positive multiple of \mathbf{x} is called an invariant measure of X . The class of irreducible recurrent chains can be further subdivided into positive and null-recurrent chains according to

whether $\sum_{i=0}^{\infty} \mathbf{x}_i \mathbf{e}^T < \infty$ or $= \infty$, with \mathbf{e} being a vector with all its components equal to one.

Vector-state Markov chains of the form (1) generalize the theory of matrix-geometric processes which was pioneered by M. Neuts [1]. As he points out in his monograph, the matrix-geometric approach is applicable to a wide class of stochastic models. The characteristic feature of these models is that the blocks within the transition probability matrix (or the generator matrix in case of continuous-time models) only depend on their relative position to the main diagonal, that is, P_{ij} only depends on $j - i$. Otherwise, Neuts' method fails.

A general strategy for solving systems of the form (1) is the following:

- (i) Truncate the system at level N , that is, consider the northwest corner truncation ${}_{(N)}P$ of P with $(N + 1) \times (N + 1)$ blocks
- (ii) Derive a stochastic matrix ${}_{(N)}\tilde{P}$ by appropriately augmenting ${}_{(N)}P$
- (iii) Solve $\mathbf{x}^{(N)} \cdot {}_{(N)}\tilde{P} = \mathbf{x}^{(N)}$.

Then $\mathbf{x}^{(N)}$ might converge to the desired solution \mathbf{x} under some appropriate conditions (a natural condition is some normalization, e.g., $x_{(0,1)}^{(N)} = 1$).

In case of (1) with $n = 1$ and P_{ij} only depending on $j - i$, *level-independent quasi-birth-death processes* arise. For $n = 1$ and arbitrary P_{ij} , the quasi-birth-death process is *level-dependent*. Several publications have dealt with the problem of computing the invariant measures for this class of stochastic processes numerically: In [3] and in [4], algorithms based on taboo probabilities were developed. In [5], a method relying on matrix continued fractions was discussed. In principle, these three algorithms differ with respect to the question how to compensate the truncation of the infinite state space which is inevitable for algorithmic procedures, that is, how to augment ${}_{(N)}P$. The major part of the algorithmic procedure is similar in all algorithms, and as pointed out in [6], for large truncation levels, it can be used without any augmentation. Since ${}_{(N)}P$ is not stochastic anymore, the last step of the above general method has to be interpreted appropriately: $\mathbf{x}^{(N)}$ shall solve $\mathbf{x}^{(N)} \cdot {}_{(N)}P = \mathbf{x}^{(N)}$ up to the first scalar equation. This simplified *matrix-analytic algorithm* becomes even more powerful when only stationary expectations have to be computed since then, the memory-efficient variant suggested in [7] can be applied.

Recently, Masuyama published a technical report [8] in which Markov chains (with a focus on continuous-time Markov chains) with a general block structure are discussed. Under somewhat restrictive conditions, it is shown that the above method (truncation, augmentation, finding a solution of the truncated and augmented system) converges to the desired invariant measure if the truncation level tends to ∞ . Since Markov chains with general block structure are discussed, transition matrices of the form (1) or its continu-

ous analogue ((20) below) can be interpreted as special cases. However, the focus in [8] is the discussion of augmentation procedures.

We will treat the problem of finding an invariant measure numerically with a different perspective: After the truncation, we perform no augmentation but find $\mathbf{x}^{(N)}$ in a similar manner as in the method for quasi-birth-death processes described in [7]. In some situations, not using an augmentation might result in requiring a higher truncation level than with an appropriate augmentation (although this is not clear). However, we discuss not only a basic algorithm for computing invariant measures but also the direct computation of stationary characteristics for which we introduce a memory-efficient implementation. Due to the memory-efficiency, our method can be applied with a very high truncation level.

In total, in some way, we combine the ideas of Neuts' original matrix-geometric approach with considerations for level-dependent quasi-birth-death processes (although the resulting transition structures are not as general as in [8]), and with memory-efficient algorithms as developed in [7] for quasi-birth-death processes.

The rest of the paper is organized as follows: We begin with a derivation of our results and methods based on taboo probabilities in Section 2 and extend the results to continuous-time Markov chains in Section 3. The resulting algorithm for computing invariant measures is presented in Section 4 whereby we also derive a memory-efficient implementation for computing stationary expectations. Section 5 is dedicated to an alternate approach. This derivation is based on generalized continued fractions or matrix continued fractions, respectively. Although these considerations do not result in new computational methods, they give us a better insight in the algebraic and numerical aspects of the approach. Finally, in Section 6, we demonstrate the applicability of our methods by briefly discussing a queueing model with batch service and server breakdowns.

2. A Derivation Based on Taboo Probabilities

According to [2, 9, 10], an invariant measure $\mathbf{x} = (x_\ell)_{\ell \in E}$ of an irreducible recurrent Markov chain X with state space E and transition probability matrix $P = (P_{ij})_{i,j \in E}$ is given by $x_{\ell_0} = 1$ and

$$x_\ell = \sum_{k=1}^{\infty} \mathbb{P}(X_k = \ell, X_{k-1}, \dots, X_1 \neq \ell_0 \mid X_0 = \ell_0), \quad (3)$$

for $\ell \neq \ell_0$, where ℓ_0 is some fixed state and $\mathbb{P}(A)$ denotes the probability of event A . The summands on the right-hand side are taboo probabilities. Irreducibility of P guarantees that the series converges for all $\ell \in E$. x_ℓ can be interpreted as the expected number of visits to state ℓ before the Markov chain returns to its initial state ℓ_0 . Note that \mathbf{x} is the unique invariant measure of P with $x_{\ell_0} = 1$.

In the present case, we choose $\ell_0 = (0, 1)$ and obtain

$$x_{(i,j)} = \sum_{k=1}^{\infty} \mathbb{P}(X_k = (i, j), X_{k-1}, \dots, X_1 \neq (0, 1) \mid X_0 = (0, 1)), \quad (4)$$

for $(i, j) \neq (0, 1)$. By $\sigma_N = \inf \{k \in \{1, 2, 3, \dots\} : X_k^{(1)} \geq N\}$, we denote the first hitting time of X of some level $\geq N$.

Following [9, 10], we have

$$\mathbf{x}_i = \lim_{N \rightarrow \infty} \frac{1}{h_{(0,1)}^{(N)}} \mathbf{h}_i^{(N)} \quad (i = 0, 1, 2, \dots), \quad (5)$$

where $\mathbf{h}^{(N)} = (\mathbf{h}_0^{(N)}, \dots, \mathbf{h}_N^{(N)})$ solves

$$\mathbf{h}^{(N)} (I_{-(N)} P) = \mathbf{e}_1. \quad (6)$$

${}_{(N)}P$ denotes the northwest corner truncation of P at level N , I is an identity matrix of appropriate size, and \mathbf{e}_1 is the vector with unity in the first position and zeroes elsewhere. The solution is unique, and with $\mathbf{x}_i^{(N)} = (1/h_{(0,1)}^{(N)}) \mathbf{h}_i^{(N)}$, we have $x_{(0,1)}^{(N)} = 1$ and

$$x_{(i,j)}^{(N)} = \sum_{k=1}^{\infty} \mathbb{P}(X_k = (i, j), X_{k-1}, \dots, X_1 \neq (0, 1), \sigma_{N+1} > k \mid X_0 = (0, 1)), \quad (7)$$

for $i \leq N$ with $(i, j) \neq (0, 1)$. Obviously, $x_{(i,j)}^{(N)}$ increases monotonically in N , and by definition, we have $\mathbf{x}_i = \lim_{N \rightarrow \infty} \mathbf{x}_i^{(N)}$.

Furthermore, we set $\tau_i = \inf \{k \in \{1, 2, 3, \dots\} : X_k^{(1)} \leq i\}$ to be the first positive hitting time of some level $\leq i$ (note that due to irreducibility and recurrence, $\tau_i < \infty$ almost surely), and we define the matrices

$$\begin{aligned} R_i &= \left(r_{i,j,j'} \right)_{\substack{1 \leq j \leq d_i \\ 1 \leq j' \leq d_{i+1}}} \\ &= \left(\sum_{k=1}^{\infty} \mathbb{P} \left(X_k = (i+1, j'), \tau_i > k \mid X_0 = (i, j) \right) \right)_{\substack{1 \leq j \leq d_i \\ 1 \leq j' \leq d_{i+1}}}, \\ R_i^{(N)} &= \left(\sum_{k=1}^{\infty} \mathbb{P} \left(X_k = (i+1, j'), \tau_i > k, \sigma_{N+1} > k \mid X_0 = (i, j) \right) \right)_{\substack{1 \leq j \leq d_i \\ 1 \leq j' \leq d_{i+1}}}. \end{aligned} \quad (8)$$

Again, the entries of R_i can be interpreted as expectations. In this case, conditioned on the initial state (i, j) , the entry at position (j, j') is the expected number of visits in $(i+1, j')$ before returning to some level $\leq i$. Similarly, the entries of $R_i^{(N)}$ are expectations for the expected number of visits in $(i+1, j')$ before returning to some level $\leq i$ or hitting some level $> N$. Obviously, $R_i^{(N)} = 0$ for $i \geq N$. Again, component-wise convergence of $R_i^{(N)} \rightarrow R_i$ is a direct consequence of the monotone convergence theorem. We are now ready to find the representation of the invariant measures which will be exploited numerically later on.

Theorem 1. *With the structure (1) of P and the notations introduced above, we have*

$$\mathbf{x}_{i+1} = \mathbf{x}_i R_i \quad (i = 0, 1, 2, \dots), \quad (9)$$

$$\mathbf{x}_{i+1}^{(N)} = \mathbf{x}_i^{(N)} R_i^{(N)} \quad (i = 0, 1, 2, \dots, N-1), \quad (10)$$

$$\mathbf{x}_0 = \mathbf{x}_0 \sum_{i=0}^n \prod_{r=0}^{i-1} R_r P_{i0}, \quad (11)$$

$$R_i = P_{i,i+1} \left(I - \sum_{j=i+1}^{i+1+n} \prod_{r=i+1}^{j-1} R_r P_{j,i+1} \right)^{-1} \quad (i = 0, 1, 2, \dots), \quad (12)$$

$$R_i^{(N)} = P_{i,i+1} \left(I - \sum_{j=i+1}^{i+1+n} \prod_{r=i+1}^{j-1} R_r^{(N)} P_{j,i+1} \right)^{-1} \quad (i = 0, 1, 2, \dots, N-1). \quad (13)$$

In (12) and (13), the inverse matrices exist. In (11), the solution of the system of linear equations is unique up to constant multiples, and additionally,

$$\mathbf{x}_0^{(N)} = \mathbf{x}_0^{(N)} \sum_{i=0}^n \prod_{r=0}^{i-1} R_r^{(N)} P_{i0} \quad (14)$$

holds up to the first scalar equation.

Proof. The interpretations of R_i are very similar to the level-independent case (see [1]) and to quasi-birth-death processes (see [4, 5]). Therefore, we do not state all proofs in detail.

(i) For $i \geq 1$, we have

$$\begin{aligned}
 x_{(i+1,j')} &= \sum_{k=2}^{\infty} \mathbb{P}\left(X_k = (i+1, j'), X_{k-1}, \dots, X_1 \neq (0, 1) \mid X_0 = (0, 1)\right) \\
 &= \sum_{k=2}^{\infty} \sum_{\ell=0}^{k-1} \mathbb{P}\left(X_k = (i+1, j'), X_{k-1}^{(1)}, \dots, X_{\ell+1}^{(1)} > i \mid X_0 = (0, 1)\right) \\
 &= \sum_{k=2}^{\infty} \sum_{\ell=1}^{k-1} \sum_{j=1}^m \mathbb{P}\left(X_k = (i+1, j'), X_{k-1}^{(1)}, \dots, X_{\ell+1}^{(1)} > i \mid X_0 = (0, 1)\right) \\
 &= \sum_{j=1}^m \sum_{k=2}^{\infty} \sum_{\ell=1}^{k-1} \mathbb{P}(X_{\ell} = (i, j), X_{\ell-1}, \dots, X_1 \neq (0, 1) \mid X_0 = (0, 1)) \\
 &\quad \cdot \mathbb{P}\left(X_k = (i+1, j'), X_{k-1}^{(1)}, \dots, X_{\ell+1}^{(1)} > i \mid X_{\ell} = (i, j)\right) \\
 &= \sum_{j=1}^m \sum_{\ell=1}^{\infty} \sum_{k=\ell+1}^{\infty} \mathbb{P}(X_{\ell} = (i, j), X_{\ell-1}, \dots, X_1 \neq (0, 1) \mid X_0 = (0, 1)) \\
 &\quad \cdot \mathbb{P}\left(X_{k-\ell} = (i+1, j'), X_{k-\ell-1}^{(1)}, \dots, X_1^{(1)} > i \mid X_0 = (i, j)\right) \\
 &= \sum_{j=1}^m \sum_{\ell=1}^{\infty} \mathbb{P}(X_{\ell} = (i, j), X_{\ell-1}, \dots, X_1 \neq (0, 1) \mid X_0 = (0, 1)) \\
 &\quad \cdot \sum_{k=1}^{\infty} \mathbb{P}\left(X_k = (i+1, j'), \tau_i > k \mid X_0 = (i, j)\right) = \sum_{j=1}^m x_{(i,j)} r_{i,j,j'}.
 \end{aligned} \tag{15}$$

This proves (9) for $i \geq 1$. For $i = 0$, we have an additional summand for $k = 1$, but on the other hand, for $k \geq 2$, the sum regarding j starts with $j = 2$. It turns out that (9) still holds. (10) is proven analogously.

(ii) Using the structure (1) of P and (9) repetitively, it becomes clear that \mathbf{x}_0 satisfies equation (11) which can be rewritten as $\mathbf{x}_0 = \mathbf{x}_0 K_0$, where

$$K_i = \sum_{j=i}^{i+n} \prod_{r=i}^{j-1} R_r P_{ji} \quad (i = 0, 1, 2, \dots). \tag{16}$$

By a calculation similar to that for proving (9), we see that the entry of K_i at position (j, j') is given by the probability

$$\begin{aligned}
 &\sum_{k=1}^{\infty} \mathbb{P}\left(X_k = (i, j'), \tau_i \geq k \mid X_0 = (i, j)\right) \\
 &= \mathbb{P}\left(\tau_i < \infty, \tau_i < \tau_{i-1}, X_{\tau_i} = (i, j') \mid X_0 = (i, j)\right),
 \end{aligned} \tag{17}$$

that the Markov chain returns to level i before hitting a level $< i$ and that this return is to (i, j') . In particular, the entries of K_0 represent the probabilities that the Markov chain eventually returns to level 0 and that this return occurs in $(0, j')$. Irreducibility and recurrence of X imply stochasticity and irreducibility (and due to K_0 being finite, recurrence) of K_0 . Hence, \mathbf{x}_0 is an invariant measure for K_0 which is uniquely defined up to a constant factor.

(iii) Alternatively, (11) can be shown by arguments similar to those leading to (9). These arguments also confirm that $\mathbf{x}_0^{(N)}$ satisfies the corresponding equation up to the first scalar equation

(iv) Additionally, define

$$J_i = \left(\sum_{k=1}^{\infty} \mathbb{P}\left(X_k = (i, j'), \tau_{i-1} > k \mid X_0 = (i, j)\right) \right)_{j, j' \in E}, \tag{18}$$

as the matrix of the expectations of the number of visits in (i, j') before hitting some level $< i$. Then considerations similar to those yielding (9) lead to $(I + J_i)(I - K_i) = (I - K_i)(I + J_i) = I$, and hence, $(I + J_i) = (I - K_i)^{-1}$ for $i \geq 1$ and $R_i = P_{i,i+1}(I + J_{i+1})$ for $i \geq 0$.

The fact that $R_i^{(N)}$ converges monotonically increasing to R_i entails that the R_i are minimal nonnegative solutions to the equations

$$R_i \left(I - \sum_{j=i+1}^{i+1+n} \prod_{r=i+1}^{j-1} R_r P_{j,i+1} \right) = P_{i,i+1}, \tag{19}$$

which corresponds to similar results in [1]. However, we will not make use of this fact in Algorithms 1 and 2 which we suggest for computing invariant measures and stationary expectations.

3. Continuous-Time Markov Chains

For continuous-time Markov chains $Y = (Y_t)_{t \geq 0}$ with block-structured generator matrix

$$Q = \begin{pmatrix} Q_{00} & Q_{01} & 0 & 0 & \dots & \dots & \dots & \dots \\ Q_{10} & Q_{11} & Q_{12} & 0 & \dots & \dots & \dots & \dots \\ \vdots & \vdots & \ddots & \ddots & & & & \\ Q_{n0} & Q_{n1} & \dots & Q_{nn} & Q_{n,n+1} & 0 & 0 & \dots \\ 0 & Q_{n+1,1} & \dots & Q_{n+1,n} & Q_{n+1,n+1} & Q_{n+1,n+2} & 0 & \dots \\ \ddots & \ddots & & \ddots & \ddots & \ddots & & \ddots \end{pmatrix}, \tag{20}$$

invariant measures satisfy $\mathbf{y}Q = 0$. For the construction of a continuous-time Markov chain from its infinitesimal generator as well as for precise definitions of properties such as regularity, recurrence, and ergodicity, the reader is referred to [11, 12]. Assuming regularity, irreducibility, and recurrence, the solution to $\mathbf{y}Q = 0$ is unique up to constant multiples and strictly positive. Furthermore, these assumptions allow constructing the embedded jump chain which is a discrete-time Markov chain $X = (X_n)_{n=0}^{\infty}$ with transition probability matrix with structure (1) and $P_{ij} = D_i^{-1}Q_{ij} + \delta_{ij}I$ where D_i is

the diagonal matrix with entries $-q_{(i,j),(i,j)}$ at position (j, j) . Irreducibility of Y implies $-q_{(i,j),(i,j)} > 0$ for all $(i, j) \in E$, and hence, D_i is indeed invertible for all $i \in \{0, 1, 2, \dots\}$. Furthermore, irreducibility and recurrence of Y imply irreducibility and recurrence of X and vice versa (note that this is not true for ergodicity). Finally, if \mathbf{x} is invariant for P , an invariant measure \mathbf{y} for Q can be constructed by setting $\mathbf{y}_i = \mathbf{x}_i D_i^{-1}$.

Renaming the matrices from Section 2 to \tilde{R}_i , we have $\mathbf{x}_i \tilde{R}_i = \mathbf{x}_{i+1}$ for the invariant measure of the embedded jump chain. Hence, we obtain

$$\mathbf{y}_{i+1} = \mathbf{x}_{i+1} D_{i+1}^{-1} = \mathbf{x}_i \tilde{R}_i D_{i+1}^{-1} = \mathbf{y}_i D_i \tilde{R}_i D_{i+1}^{-1} = \mathbf{y}_i R_i, \quad (21)$$

with $R_i = D_i \tilde{R}_i D_{i+1}^{-1}$. From (11), we obtain

$$\begin{aligned} \mathbf{y}_0 &= \mathbf{x}_0 D_0^{-1} = \mathbf{x}_0 \sum_{i=0}^n \prod_{r=0}^{i-1} \tilde{R}_r P_{i0} D_0^{-1} \\ &= \mathbf{y}_0 D_0 \sum_{i=0}^n \prod_{r=0}^{i-1} D_r^{-1} R_r D_{r+1} (D_i^{-1} Q_{i0} + \delta_{i0} I) D_0^{-1} \\ &= \mathbf{y}_0 \sum_{i=0}^n \prod_{r=0}^{i-1} R_r Q_{i0} + \mathbf{y}_0. \end{aligned} \quad (22)$$

Taking into account that \tilde{R}_i satisfies (12), we arrive at

$$\begin{aligned} R_i &= D_i \tilde{R}_i D_{i+1}^{-1} = D_i P_{i,i+1} \left(I - \sum_{j=i+1}^{i+1+n} \prod_{r=i+1}^{j-1} \tilde{R}_r P_{j,i+1} \right)^{-1} D_{i+1}^{-1} \\ &= D_i D_i^{-1} Q_{i,i+1} \left(D_{i+1} - D_{i+1} \sum_{j=i+1}^{i+1+n} \prod_{r=i+1}^{j-1} D_r^{-1} R_r D_{r+1} \right. \\ &\quad \left. \cdot (D_j^{-1} Q_{j,i+1} + \delta_{j,i+1} I) \right)^{-1} \\ &= -Q_{i,i+1} \left(\sum_{j=i+1}^{i+1+n} \prod_{r=i+1}^{j-1} R_r Q_{j,i+1} \right)^{-1} \quad (i = 0, 1, 2, \dots). \end{aligned} \quad (23)$$

Since the j' th diagonal entry of D_{i+1}^{-1} can be interpreted as the mean sojourn time in $(i+1, j')$, the entries of $\tilde{R}_i D_{i+1}^{-1}$ are the expected time intervals spent in state $(i+1, j')$ before returning to some level $\leq i$, conditioned on starting in (i, j) . Therefore, the entries of R_i represent the proportion of the expected times spent in $(i+1, j')$ before returning to some level $\leq i$, and the time spent in the initial state (i, j) .

Summarizing, we get

Theorem 2. *With the above assumptions on Q , we have*

$$\mathbf{y}_{i+1} = \mathbf{y}_i R_i \quad (i = 0, 1, 2, \dots), \quad (24)$$

$$\mathbf{y}_{i+1}^{(N)} = \mathbf{y}_i^{(N)} R_i^{(N)} \quad (i = 0, 1, 2, \dots, N-1), \quad (25)$$

$$0 = \mathbf{y}_0 \sum_{i=0}^n \prod_{r=0}^{i-1} R_r Q_{i0}, \quad (26)$$

where the solution of the system of linear equations in (26) is unique up to constant multiples. The R_i 's satisfy

$$R_i = -Q_{i,i+1} \left(\sum_{j=i+1}^{i+1+n} \prod_{r=i+1}^{j-1} R_r Q_{j,i+1} \right)^{-1} \quad (i = 0, 1, 2, \dots), \quad (27)$$

and can be uniquely constructed by $R_i = \lim_{N \rightarrow \infty} R_i^{(N)}$ for $i = 0, 1, 2, \dots$, where $R_i^{(N)} = 0$ for $i \geq N$ and

$$R_i^{(N)} = -Q_{i,i+1} \left(\sum_{j=i+1}^{i+1+n} \prod_{r=i+1}^{j-1} R_r^{(N)} Q_{j,i+1} \right)^{-1} \quad (i = 0, 1, 2, \dots, N-1). \quad (28)$$

$\mathbf{y}_0^{(N)}$ satisfies

$$0 = \mathbf{y}_0^{(N)} \sum_{i=0}^n \prod_{r=0}^{i-1} R_r^{(N)} Q_{i0}, \quad (29)$$

up to the first scalar equation.

4. Algorithms

Next, we demonstrate how to use the results numerically. We focus on continuous-time Markov chains since the methods for discrete-time Markov chains follow an analogous scheme.

Basically, we have to use (28) for computing approximations to R_{N-1}, \dots, R_0 , then we determine an approximation for \mathbf{y}_0 by solving (26) up to the first scalar equation, and finally, we use (24) for computing approximations to $\mathbf{y}_1, \dots, \mathbf{y}_N$. In order to save operations, we suggest the following algorithm for computing an approximation to the invariant distribution.

In case of positive recurrence, an invariant distribution can be obtained afterwards by renormalizing \mathbf{y} . We give some remarks:

- (i) Algorithm 3 in [8] is quite similar. While our algorithm relies on the nonstochastic matrix $P = (P_{in})_{i,n=1}^N$ or the nonconservative matrix $Q = (Q_{in})_{i,n=1}^N$, respectively, in [8], an augmentation is performed first in order to obtain a stochastic or conservative matrix. Note that Theorem 1 guarantees that for $N \rightarrow \infty$, we have convergence of the approximations $\mathbf{x}^{(N)}$ or $\mathbf{y}^{(N)}$ to the desired invariant measure without any augmentation. The only use of augmentations is that the truncation error might be decreased a bit for small values of the truncation level N . Instead of considering an augmentation, we suggest to choose N very large. If such a choice results in computational problems, we suggest to use the memory-efficient algorithm


```

Choose  $N$  large
Initialize  $H_j = -Q_{Nj}$  for  $j = N - n, \dots, N$ 
for  $i = N - 1, N - 2, \dots, 0$  do
    Compute  $R_i = Q_{i,i+1}H_{i+1}^{-1}$ 
    Free memory space for  $H_{i+1}$ 
    Update  $H_j = -Q_{ij} + R_iH_j$  for  $j = \max\{i - n + 1, 0\}, \dots, i$ 
    Initialize  $H_{i-n} = -Q_{i,i-n}$  if  $i - n > 0$ 
end for
Set  $y_{(0,1)} = 1$  and determine all other entries of  $y_0$  by solving  $y_0H_0$  up to the first scalar equation.
For  $i = 0, 1, \dots, N - 1$ , compute  $y_{i+1} = y_iR_i$ 
return  $y$ 

```

ALGORITHM 1: Computation of an invariant measure.

for computing stationary expectations (see below). Furthermore, note that the results in [8] were derived under different conditions: While $P_{in} \neq 0$ is allowed for all $i \leq n + 1$ (lower-block Hessenberg matrices), some restrictive conditions (Assumptions 2.1 and 2.2 in [8]) are assumed

- (ii) A major problem is the choice of the truncation level N . There are some approaches concerning a priori estimates of the truncation error (e.g., [13]), but still the problem is not completely solved. Instead of using such error estimates, alternatively, as in [8], we might increase the truncation level iteratively, that is, we start with some small N and increase N by 1 in each step until $\|y^{(N)} - y^{(N-1)}\| < \varepsilon$ for some fixed norm $\|\cdot\|$ and some fixed small $\varepsilon > 0$. Note that for each new value of N , we have to restart Algorithm 1. Furthermore, note that a small difference $\|y^{(N)} - y^{(N-1)}\|$ does not imply that $\|y^{(N)} - y\|$ is small where y is the true invariant measure
- (iii) In many applications, the state space is finite with a transition structure corresponding to the northwest corner of (1) or (20), respectively. Then Algorithm 1 is exact up to numerical deviations
- (iv) Instead of computing the inverse matrix H_{i+1}^{-1} first and then perform the matrix multiplication for determining $R_i = Q_{i,i+1}H_{i+1}^{-1}$, it is recommended to solve the system $R_iH_{i+1} = Q_{i,i+1}$ directly
- (v) The matrices Q_{ij} can be generated when they are used for computation. Hence, there is no need to store these matrices
- (vi) We have to store about N matrices at the same time. If the dimensions d_i of the matrices are constant, say $d_i = d$, the memory requirement is approximately Nd^2 real numbers
- (vii) Concerning the computational effort, we have to solve approximately N linear systems of equations of the form $AX = B$ where the coefficients A and B and the solution X are matrices of dimension

$d \times d$ (for constant $d_i = d$). In addition, we have to perform $N \cdot n$ matrix multiplications. The computational effort is approximately $\gamma_1 Nd^3 + \gamma_2 Nnd^\alpha$ with some constants γ_1, γ_2 and $\alpha \in (2, 3]$ (the exponent α and the constants γ_1, γ_2 depend on the method of choice, we have $\alpha = 3$ for “conventional matrix multiplication” and $\alpha = 2.807$ for Strassen’s algorithm, see [14])

- (viii) For $n = 1$, Algorithm 1 more or less coincides with the methods proposed in [3–5]. One might expect that also in general, a discrete-time Markov chain with transition probability matrix (1) and bandwidth > 3 or a continuous-time Markov chain with generator matrix (20) and bandwidth > 3 , respectively, may be reduced to the block-tridiagonal case by appropriately rearranging the states, i.e., define a new level 0 as the union of the original levels $0, \dots, n - 1$ and define a new level 1 as the union of the original levels $n, \dots, 2n - 1, \dots$. But the increased block size heavily impacts the performance of the algorithm. If we apply the same truncation procedure, the truncation level for the modified matrix structure is $\approx N/n$, and for $d_i = d$, all matrices involved have the dimension $(nd) \times (nd)$. This results in a computational effort of approximately $\gamma_1 (N/n)(nd)^3 + \gamma_2 (N/n)(nd)^\alpha = \gamma_1 Nn^2d^3 + \gamma_2 Nn^{\alpha-1}d^\alpha$. Since $\alpha > 2$, the computational effort for the transformed system becomes significantly larger than the effort for the original system. Similarly, the memory requirement of Nd^2 is replaced by $(N/n) \cdot (nd)^2 = Nnd^2$. For these reasons, one should discard this supposedly simple “trick” (compare ([15], pp.~73ff)).

In applications, often only stationary expectations have to be considered: Consider the case of positive recurrence, and denote by π the invariant distribution. Then the ergodic theorem for discrete-time Markov chains guarantees that

$$\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=0}^{K-1} f(X_k) = \pi f = \frac{\mathbf{x}f}{\mathbf{x}\mathbf{1}}, \quad (30)$$

```

Choose  $N$  large
Initialize  $H_j = -Q_{Nj}$  for  $j = N - n, \dots, N$  and  $Z = f(N)$ 
for  $i = N - 1, N - 2, \dots, 0$  do
    Update  $R = Q_{i,i+1} H_{i+1}^{-1}$ 
    Update  $Z = f(i) + RZ$ 
    Free memory space for  $H_{i+1}$ 
    Update  $H_j = -Q_{ij} + RH_j$  for  $j = \max\{i - n + 1, 0\}, \dots, i$ 
    Initialize  $H_{i-n} = -Q_{i,i-n}$  if  $i - n \geq 0$ 
end for
Set  $y_{(0,1)} = 1$  and determine all other entries of  $\mathbf{y}_0$  by solving  $\mathbf{y}_0 H_0$  up to the first scalar equation
Update  $Z = \mathbf{y}_0 Z$  and renormalize  $Z = (z_1, \dots, z_{L+1})$  such that  $z_{L+1} = 1$ 
return  $Z$ 

```

ALGORITHM 2: Computation of stationary expectations.

for any function $f : E \rightarrow \mathbb{R}$ for which $\pi f := \sum_{\ell \in E} \pi_\ell f(\ell)$ converges absolutely. Again, by **1**, we refer to a vector of appropriate dimension consisting of 1s. Analogously, for continuous-time Markov chains, we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(X_t) dt = \pi f = \frac{\mathbf{y}f}{\mathbf{y}\mathbf{1}}, \quad (31)$$

if πf converges absolutely. Now, let $f = (f_1, \dots, f_L, f_{L+1}) : E \rightarrow \mathbb{R}^{L+1}$ be a vector valued function with $f_{L+1}(\ell) = 1$ for all $\ell \in E$. Then for continuous-time Markov chains, we have $\pi f_\ell = \mathbf{y}f_\ell / \mathbf{y}f_{L+1}$ for $\ell = 1, \dots, L$. Hence, computing $\mathbf{y}f$ allows to determine the stationary characteristics πf_ℓ which coincide with the long-run costs or rewards measured by f_ℓ .

When computing $\mathbf{y}f$, we can apply a Horner-type scheme which allows us not to store the matrices R_i and therefore save quite a lot memory requirement. This idea is very similar to that in [7]. With

$$f(i) = \begin{pmatrix} f_1((i, 1)) & \cdots & f_{L+1}((i, 1)) \\ \vdots & \ddots & \vdots \\ f_1((i, d_i)) & \cdots & f_{L+1}((i, d_i)) \end{pmatrix}, \quad (32)$$

we write

$$\begin{aligned} \mathbf{y}f &\approx \sum_{j=0}^N \mathbf{y}_j^{(N)} f(j) = \mathbf{y}_0^{(N)} \sum_{j=0}^n \prod_{r=i}^{j-1} R_r^{(N)} f(j) \\ &= \mathbf{y}_0^{(N)} \left(f(0) + R_0^{(N)} \left(f(1) + R_1^{(N)} \left(f(2) \right. \right. \right. \\ &\quad \left. \left. \left. + R_2^{(N)} \left(\cdots + R_{N-2}^{(N)} \left(f(N-1) + R_{N-1}^{(N)} f(N) \right) \right) \right) \right) \right) \\ &= \mathbf{y}_0^{(N)} Z_0^{(N)}, \end{aligned} \quad (33)$$

where

$$\begin{aligned} Z_i^{(N)} &= \sum_{j=i}^n \prod_{r=i}^{j-1} R_r^{(N)} f(j) = f(i) + R_i^{(N)} \left(f(i+1) \right. \\ &\quad \left. + R_{i+1}^{(N)} \left(\cdots + R_{N-2}^{(N)} \left(f(N-1) + R_{N-1}^{(N)} f(N) \right) \right) \right). \end{aligned} \quad (34)$$

Obviously, we have $Z_N^{(N)} = f(N)$ and

$$Z_i^{(N)} = f(i) + R_i^{(N)} Z_{i+1}^{(N)} \quad (i = 0, \dots, N-1). \quad (35)$$

The matrices $R_i^{(N)}$ are computed by a backward scheme, and (35) is a backward scheme, too. Since we are only interested in computing $Z_0^{(N)}$, there is no need to store the matrices $R_i^{(N)}$ or $Z_i^{(N)}$. We suggest

Again, we give some comments:

- (i) The problem of finding an appropriate truncation level N to keep the error in prescribed limits is still to be solved. In [16], a generalization of the idea from [13] is discussed, and $\sum_{i=0}^N \mathbf{y}_i f(i)$ and $\sum_{i=0}^{\infty} \mathbf{y}_i f(i)$ are compared. Unfortunately, the method is quite cumbersome to apply, especially because it would be more desirable to compare $\sum_{i=0}^{\infty} \mathbf{y}_i f(i)$ with $\sum_{i=0}^N \mathbf{y}_i^{(N)} f(i)$. Again, an alternative consists in iteratively increasing N until $\sum_{i=0}^N \mathbf{y}_i^{(N)} f(i) - \sum_{i=0}^{N-1} \mathbf{y}_i^{(N-1)} f(i)$ becomes small. Note that again, this method is somehow problematic since the value of this difference does not tell anything about the true error $\sum_{i=0}^N \mathbf{y}_i^{(N)} f(i) - \sum_{i=0}^{\infty} \mathbf{y}_i f(i)$.
- (ii) For finite state spaces with the transition structure under consideration, Algorithm 2 is exact up to numerical deviations
- (iii) Of course, it is still recommended to solve the system $RH_{i+1} = Q_{i,i+1}$ directly, and still, the matrices Q_{ij} should be generated when they are needed

- (iv) In total, we have to store the matrices H_i and Z . Since we need n matrices H_i at the same time, the total amount of memory requirement is $O(nd^2)$ if the matrix dimensions $d_i = d$ are constant. In particular, the memory requirement does not depend directly on the truncation level N
- (v) As pointed out above, in ([8], section 3.2), lower-block Hessenberg matrices (block $GI/M/1$ -type matrices) are considered, that is, $P_{ij} = 0$ for $j > i + 1$, but $P_{ij} \neq 0$ is allowed for all $j < i$. All our theoretical considerations carry over (we only have to replace n by ∞ or n by N at some points). Unfortunately, when applying Algorithm 2 to such block matrices, we still have to store N matrices, and hence, it is not memory-efficient anymore. Since the memory-efficient implementation makes the theoretical considerations applicable to a wide range of stochastic models, we have restricted ourselves to considering band-block-structured matrices
- (vi) The considerations concerning the computational effort are similar to those for Algorithm 1. In particular, we do not recommend to reduce the structure to the case $n = 1$
- (vii) Note that often, “natural” state spaces are multi-dimensional. However, if there is a partition $E = E_0 \cup E_1 \cup E_2 \cup \dots$ into finite levels E_i such that transitions with positive probability/transition rate increase the level E_i at most by 1 and decrease it at most by n , an appropriate renumbering of the states results in a structure given by (1) or (20). We refer to the end of Section 6 for an example and to [7, 13, 17] for the corresponding discussion concerning quasi-birth-death processes (that is, $n = 1$).

5. Relationship to Matrix Continued Fractions

This section is devoted to the interdependence between Markov chains and matrix continued fractions. The results obtained are of importance not only from an inner mathematical but also from a numerical point of view. Matrix continued fractions are generalizations of ordinary continued fractions and generalized continued fractions (or n -fractions, see [18, 19]). We will show that with each transition matrix P of the form (1), one can associate a matrix continued fraction (or a matrix-driven Jacobi-Perron algorithm) being convergent if X is irreducible and recurrent. In addition, it is proven that its approximants and its limits coincide with the coefficients of the recurrence relations (10) and (9), respectively.

A milestone in pure mathematics has been the arithmetic characterization of quadratic irrationalities which has been discovered by J. L. Lagrange (see [20]). Lagrange’s theorem states that the expansion of a real number as a simple continued fraction is periodic if and only if it is the real root of a quadratic equation in one variable with rational coefficients. Since then, mathematicians like C. G. J. Jacobi [21],

O. Perron [22], P. Bachmann [23], L. Bernstein [24], and F. Schweiger [25] tried to expand Lagrange’s approach to higher degree irrationalities. G. C. F. Jacobi and O. Perron developed a tool for approximating real algebraic irrationalities of degree $n \geq 3$. To the honor of Jacobi and Perron, this algorithm has been named as the Jacobi-Perron algorithm. Although the Jacobi-Perron algorithm does not lead to the desired results, it has been successfully addressed to a lots of other unsolved problems in approximation theory and numerical analysis. In particular, in connection with modern computer technology, the Jacobi-Perron algorithm turns out to be an efficient tool for identifying and calculating subdominant solutions of higher order linear vector recursions.

In the sequel, we will show that the schemes (10) and (13) are equivalent to a specific matrix continued fraction or a matrix-driven Jacobi-Perron algorithm associated with the transition probability matrix (1). For this purpose, we assume that $d_i = d$ for all $i = 0, 1, 2, \dots$ and that $P_{i,i+1}$ is invertible for all $i = 0, 1, 2, \dots$. Define

$$A_{i+n-k}^{(k)} = \begin{cases} -P_{i+n-k,i} P_{i-1,i}^{-1} & \text{for } k = 0, 1, \dots, n-1; i = 1, 2, \dots, \\ -(P_{ii} - I) P_{i-1,i}^{-1} & \text{for } k = n; i = 1, 2, \dots, \end{cases} \quad (36)$$

where I denotes the unit matrix. Then (36) and (2) yield the $(n+1)$ th-order homogeneous linear vector recursion

$$\mathbf{x}_{i-1} = \mathbf{x}_i A_i^{(n)} + \mathbf{x}_{i+1} A_{i+1}^{(n-1)} + \dots + \mathbf{x}_{i+n} A_{i+n}^{(0)} \quad (i = 1, 2, \dots), \quad (37)$$

subject to the boundary condition

$$\mathbf{x}_0(P_{00} - I) + \mathbf{x}_1 P_{10} + \dots + \mathbf{x}_n P_{n0} = 0. \quad (38)$$

In this perspective, the problem at hand is to find the “right” initial values, say $\mathbf{x}_1, \mathbf{x}_{l+1}, \dots, \mathbf{x}_{l+n}$, l sufficiently large, such that $\mathbf{x}_i > 0$ for all i . In order to find a solution to this problem based on generalized continued fractions, we generalize and combine the results from [5, 26].

Remember that

$$\mathbf{x}_i = \lim_{N \rightarrow \infty} {}_{(N)}h_{(0,1)(N)}^{-1} \mathbf{h}_i \quad (i = 0, 1, 2, \dots), \quad (39)$$

where ${}_{(N)}\mathbf{h} = ({}_{(N)}\mathbf{h}_0, \dots, {}_{(N)}\mathbf{h}_N)$ solves the truncated system (6). (36) and (6) lead to the finite difference scheme

$$\begin{aligned} {}_{(N)}\mathbf{h}_{i-1} &= {}_{(N)}\mathbf{h}_i A_i^{(n)} + {}_{(N)}\mathbf{h}_{i+1} A_{i+1}^{(n-1)} + \dots \\ &\quad + {}_{(N)}\mathbf{h}_{i+n} A_{i+n}^{(0)} \quad (i = 1, 2, \dots, N), \end{aligned} \quad (40)$$

coupled to the boundary conditions:

$$\mathbf{e}_1 = {}_{(N)}\mathbf{h}_0(P_{00} - I) + {}_{(N)}\mathbf{h}_1 P_{10} + \dots + {}_{(N)}\mathbf{h}_n P_{n0}, \quad (41)$$

$${}_{(N)}\mathbf{h}_{N+1} = {}_{(N)}\mathbf{h}_{N+2} = \dots = {}_{(N)}\mathbf{h}_{N+n} = 0, \quad (42)$$

where \mathbf{e}_1 is a $1 \times m$ vector with unity in the first position and zeroes elsewhere. For simplification we transform the system (40) into a first order scheme of the form

$$({}_N)\mathbf{h}(i-1) = ({}_N)\mathbf{h}(i) \cdot A(i) \quad (i = 1, \dots, N), \quad (43)$$

by putting

$$({}_N)\mathbf{h}(i) = \left(({}_N)\mathbf{h}_{i,(N)}\mathbf{h}_{i+1}, \dots, ({}_N)\mathbf{h}_{i+n} \right) \quad (i = 0, 1, 2, \dots),$$

$$A(i) = \begin{bmatrix} A_i^{(n)} & I & 0 & \dots & 0 & 0 \\ A_{i+1}^{(n-1)} & 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ A_{i+n-1}^{(1)} & 0 & 0 & \dots & 0 & I \\ A_{i+n}^{(0)} & 0 & 0 & \dots & 0 & 0 \end{bmatrix} \quad (i = 1, 2, \dots). \quad (44)$$

Now, let $\ell \leq N$ be fixed. By repeated substitutions, we obtain from (43)

$$({}_N)\mathbf{h}(\ell) = ({}_N)\mathbf{h}(N) \Phi_\ell(N - \ell), \quad (45)$$

with

$$\Phi_\ell(N - \ell) = A(N)A(N-1) \dots A(\ell+1). \quad (46)$$

For the remainder, it will be convenient to partition $\Phi_\ell(i)$ as

$$\Phi_\ell(i) = \begin{bmatrix} \Phi_{\ell,i}^{00} & \Phi_{\ell,i}^{01} & \dots & \Phi_{\ell,i}^{0n} \\ \Phi_{\ell,i}^{10} & \Phi_{\ell,i}^{11} & \dots & \Phi_{\ell,i}^{1n} \\ \vdots & \vdots & & \vdots \\ \Phi_{\ell,i}^{n0} & \Phi_{\ell,i}^{n1} & \dots & \Phi_{\ell,i}^{nn} \end{bmatrix} \quad (i = 0, 1, 2, \dots), \quad (47)$$

where the matrices $\Phi_{\ell,i}^{rs}$ are of dimension $d \times d$. In view of the boundary condition (42), we infer from (45) that

$$({}_N)\mathbf{h}_{\ell+k} = ({}_N)\mathbf{h}_N \Phi_{\ell,N-\ell}^{0k} \quad (k = 0, 1, \dots, n). \quad (48)$$

For computing the $d \times d$ -matrices $\Phi_{\ell,N-\ell}^{0k}$, the following result reduces the computational effort.

Theorem 3. Let $A_{i+n-k+\ell}^{(k)}$, $k = 0, 1, \dots, n$; $i = 1, 2, \dots$, be defined as in (36) and let, for convenience,

$$A_{i+\ell}^{(k)} = \begin{cases} I & \text{for } k = 0; i = 1, \dots, n; \\ 0 & \text{for } k = 1, \dots, n-1; i = 1, \dots, n-k. \end{cases} \quad (49)$$

Then

$$\Phi_{\ell,i}^{0k} = A_{i+\ell}^{(n)} \Phi_{\ell,i-1}^{0k} + A_{i+\ell}^{(n-1)} \Phi_{\ell,i-2}^{0k} + \dots + A_{i+\ell}^{(0)} \Phi_{\ell,i-n-1}^{0k} \quad (k = 0, 1, \dots, n; i = 1, 2, \dots), \quad (50)$$

where

$$\Phi_{\ell,-j}^{00} = \begin{cases} I & \text{for } j = 0, \\ 0 & \text{for } j = 1, \dots, n \end{cases}$$

$$\Phi_{\ell,-j}^{0k} = \begin{cases} I & \text{for } j = n-k+1 \\ 0 & \text{elsewhere} \end{cases} \quad (k = 1, \dots, n; j = 0, 1, \dots, n) \quad (51)$$

Proof. From (46), it is clear that the matrices $\Phi_\ell(i)$ can be recursively computed by

$$\Phi_\ell(i) = A(i+\ell) \Phi_\ell(i-1) \quad (i = 1, 2, \dots), \quad (52)$$

with initial condition

$$\Phi_\ell(0) = I. \quad (53)$$

Rewriting (52) in expanded form we get

$$\left. \begin{aligned} \Phi_{\ell,i}^{00} &= A_{i+\ell}^{(n)} \Phi_{\ell,i-1}^{00} + \Phi_{\ell,i-1}^{10} \\ \Phi_{\ell,i}^{01} &= A_{i+\ell}^{(n)} \Phi_{\ell,i-1}^{01} + \Phi_{\ell,i-1}^{11} \\ &\vdots \\ \Phi_{\ell,i}^{10} &= A_{i+\ell+1}^{(n-1)} \Phi_{\ell,i-1}^{00} + \Phi_{\ell,i-1}^{20} \\ \Phi_{\ell,i}^{11} &= A_{i+\ell+1}^{(n-1)} \Phi_{\ell,i-1}^{01} + \Phi_{\ell,i-1}^{21} \\ &\vdots \\ \Phi_{\ell,i}^{n0} &= A_{i+\ell+n}^{(0)} \Phi_{\ell,i-1}^{00} \\ \Phi_{\ell,i}^{n1} &= A_{i+\ell+n}^{(0)} \Phi_{\ell,i-1}^{01} \\ &\vdots \end{aligned} \right\} \quad (i = 1, 2, \dots). \quad (54)$$

By repeated substitutions, it is readily seen that the matrices $\Phi_{\ell,i}^{0k}$ satisfy the recurrence relation (50) at least for $i = n+1, n+2, \dots$. To proceed in this manner we assume, for the moment, that the recurrence relation (50) may be continued for $i = 0, -1, \dots, -n+1$, without affecting the initial condition (53) and the scheme (54). Since the matrices $A(\ell), \dots, A(-n+1+\ell)$ do not appear in formula (46), they can be chosen arbitrarily. By putting

$$\begin{aligned}
A(-j+\ell) &= \begin{bmatrix} A_{-j+\ell}^{(n)} & I & 0 & \cdots & 0 \\ A_{-j+1+\ell}^{(n-1)} & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ A_{-j+n-1+\ell}^{(1)} & 0 & 0 & \cdots & I \\ A_{-j+n+\ell}^{(0)} & 0 & 0 & \cdots & 0 \end{bmatrix} = M \\
&= \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & I \\ I & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (j=0, \dots, n-1), \\
\end{aligned} \tag{55}$$

and $\Phi_\ell(-n) = M$ we get

$$\Phi_\ell(-n+1) = M^2 = \begin{bmatrix} 0 & 0 & I & 0 & \cdots & 0 \\ 0 & 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ I & 0 & 0 & 0 & \cdots & 0 \\ 0 & I & 0 & 0 & \cdots & 0 \end{bmatrix}, \dots, \Phi_\ell(-1)$$

$$= M^n = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & I \\ I & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 & 0 \\ 0 & 0 & \cdots & 0 & I & 0 \end{bmatrix},$$

(56)

which explains (49) and (51). Since $\Phi_\ell(0) = M^{n+1} = I$, (55) becomes compatible with (52) and (53), which proves our result for $i = 1, \dots, n$.

To make (48) congruent with (9) and (10), respectively, we first prove that $\det(\Phi_{i-1, N-(i-1)}^{00}) \neq 0$ for almost all N under a mild condition on P .

Consider the recurrence relation (40) for $i = \ell + 1, \ell + 2, \dots, N$. Assembling the unknowns ${}_{(N)}\mathbf{h}_\ell, {}_{(N)}\mathbf{h}_{\ell+1}, \dots, {}_{(N)}\mathbf{h}_N$ into an $1 \times m \cdot (N - \ell + 1)$ vector and setting the boundaries ${}_{(N)}\mathbf{h}_N = \nu, {}_{(N)}\mathbf{h}_{N+1} = {}_{(N)}\mathbf{h}_{N+2} = \dots = {}_{(N)}\mathbf{h}_{N+n} = 0$, (40) permits a presentation of the form

$$\left({}_{(N)}\mathbf{h}_\ell, {}_{(N)}\mathbf{h}_{\ell+1}, \dots, {}_{(N)}\mathbf{h}_N \right) W = [0, \dots, 0, \nu], \tag{57}$$

where

$$W = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \left(\begin{array}{cccccccc|c} P_{\ell, \ell+1} & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ P_{\ell+1, \ell+1} & -I & P_{\ell+1, \ell+2} & 0 & \cdots & \cdots & \cdots & 0 \\ P_{\ell+2, \ell+1} & P_{\ell+2, \ell+2} & -I & P_{\ell+2, \ell+3} & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{\ell+n+1, \ell+1} & P_{\ell+n+1, \ell+2} & & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & P_{\ell+n+2, \ell+2} & & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & P_{N-1, N-n} & \cdots & P_{N-1, N} & 0 \\ 0 & 0 & \cdots & \cdots & P_{N, N-n} & \cdots & P_{N, N} - I & 0 \end{array} \right) \tag{58}$$

Since we assumed $\det(P_{i, i+1}) \neq 0$ for all i , W is a regular matrix. For its inverse, we choose the partition

$$W^{-1} = \begin{pmatrix} X & Y \\ U & V \end{pmatrix}, \tag{59}$$

where U is a square matrix of order d . From (48), we know that

$${}_{(N)}\mathbf{h}_\ell = {}_{(N)}\mathbf{h}_N \Phi_{\ell, N-\ell}^{00}. \tag{60}$$

Solving (57) successively with v replaced by $(1, 0, \dots, 0)$, $(0, 1, \dots, 0)$, ..., $(0, \dots, 0, 1)$, we get

$$\Phi_{\ell, N-\ell}^{00} = U. \quad (61)$$

The matrix U is called the algebraic complement of the matrix C . A theorem which is due to Jacobi (see ([27], pp. 123-124, formula (4.45))) states that $\det U \neq 0$ iff $\det C \neq 0$. Since we assumed $\det(P_{N, N+1}) \neq 0$, C is weakly diagonally dominant. To guarantee that C is regular, it is therefore sufficient to assume that C is irreducible (see ([28], p. 531, Theorem 3)).

From (48), we conclude that

$${}_{(N)}\mathbf{h}_{\ell+k} = {}_{(N)}\mathbf{h}_N \Phi_{\ell, N-\ell}^{0k} \quad (k = 1, \dots, n) \text{ and } {}_{(N)}\mathbf{h}_N = {}_{(N)}\mathbf{h}_\ell (\Phi_{\ell, N-\ell}^{00})^{-1}, \quad (62)$$

implying

$${}_{(N)}\mathbf{h}_{\ell+k} = {}_{(N)}\mathbf{h}_\ell R_{\ell, \ell+k}^{(N)} \quad (k = 1, \dots, n; \ell = 0, 1, \dots, N), \quad (63)$$

where

$$R_{\ell, \ell+k}^{(N)} = (\Phi_{\ell, N-\ell}^{00})^{-1} \Phi_{\ell, N-\ell}^{0k} \quad (k = 1, \dots, n; \ell = 0, 1, \dots, N). \quad (64)$$

From (63) and (41), we obtain the decoupled recursion

$${}_{(N)}\mathbf{h}_{\ell+1} = {}_{(N)}\mathbf{h}_\ell R_{\ell, \ell+1}^{(N)} \quad (\ell = 0, 1, \dots, N), \quad (65)$$

and the boundary condition

$${}_{(N)}\mathbf{h}_0 (P_{00} - I + R_{01}^{(N)} P_{10} + \dots + R_{0N}^{(N)} P_{n0}) = \mathbf{e}_1, \quad (66)$$

which enable us to compute ${}_{(N)}\mathbf{h}_\ell$ for $\ell = 0, 1, \dots, N$. In view of the representations (49), (50), and (51), we make use of the notation

$$\begin{bmatrix} R_{\ell, \ell+1}^{(N)} \\ R_{\ell, \ell+2}^{(N)} \\ \vdots \\ R_{\ell, \ell+n}^{(N)} \end{bmatrix} = \begin{bmatrix} A_{\ell+1}^{(0)} & A_{\ell+2}^{(0)} & A_{\ell+3}^{(0)} & \cdots & A_N^{(0)} \\ A_{\ell+1}^{(1)} & A_{\ell+2}^{(1)} & A_{\ell+3}^{(1)} & \cdots & A_N^{(1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{\ell+1}^{(n-1)} & A_{\ell+2}^{(n-1)} & A_{\ell+3}^{(n-1)} & \cdots & A_N^{(n-1)} \\ A_{\ell+1}^{(n)} & A_{\ell+2}^{(n)} & A_{\ell+3}^{(n)} & \cdots & A_N^{(n)} \end{bmatrix}$$

$$= \begin{bmatrix} I & I & \cdots & I & A_{\ell+n+1}^{(0)} & \cdots & A_N^{(0)} \\ 0 & 0 & \cdots & A_{\ell+n}^{(1)} & A_{\ell+n+1}^{(1)} & \cdots & A_N^{(1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & A_{\ell+2}^{(n-1)} & \cdots & A_{\ell+n}^{(n-1)} & A_{\ell+n+1}^{(n-1)} & \cdots & A_N^{(n-1)} \\ A_{\ell+1}^{(n)} & A_{\ell+2}^{(n)} & \cdots & A_{\ell+n}^{(n)} & A_{\ell+n+1}^{(n)} & \cdots & A_N^{(n)} \end{bmatrix}. \quad (67)$$

(67) can be interpreted as the N th convergent of the generalized matrix continued fraction (GCF) or the matrix-driven Jacobi-Perron algorithm associated with the limiting process

$$R_{\ell, \ell+k} = \lim_{N \rightarrow \infty} R_{\ell, \ell+k}^{(N)} \quad (k = 1, \dots, n), \quad (68)$$

compare with [22, 29, 30]. If the limits (68) do exist, we obtain

$$\begin{aligned} \mathbf{x}_{\ell+1} &= \mathbf{x}_\ell R_{\ell, \ell+1}, \quad (\ell = 0, 1, 2, \dots), \\ \mathbf{x}_0 (P_{00} - I + R_{01} P_{10} + \dots + R_{0n} P_{n0}) &= 0, \end{aligned} \quad (69)$$

as a consequence of (65), (39), and (38).

For the calculation of $R_{\ell, \ell+k}^{(N)}$ and $R_{\ell, \ell+k}$, we recommend the following algorithms:

- Forward algorithm (FA)*: this is nothing else than a straightforward application of the relations (64), (49), (50), and (51).
- Backward algorithm (BA)*: since a matrix-driven Jacobi-Perron algorithm may be interpreted as a special case of a matrix continued fraction, its N th convergent $R_{\ell, \ell+k}^{(N)} = (\Phi_{\ell, N-\ell}^{00})^{-1} \Phi_{\ell, N-\ell}^{0k}$, $k = 1, \dots, n$, can be calculated from a linear fractional transformation (see ([29], Theorem 2.2)). Restricted to the matrix-driven Jacobi-Perron algorithm, this algorithm runs as follows. For N sufficiently large, define

$$R_{N+1, N+1+k}^{(N)} = 0 \quad (k = 1, \dots, n). \quad (70)$$

Then

$$R_{\ell, \ell+1}^{(N)} = \left(B_\ell + R_{\ell+1, \ell+1+n}^{(N)} \right)^{-1} A_\ell^{(0)} \quad (\ell = N, N-1, \dots, 0), \quad (71)$$

$$R_{\ell, \ell+k}^{(N)} = \left(B_\ell + R_{\ell+1, \ell+1+n}^{(N)} \right)^{-1} \left(A_\ell^{(k-1)} + R_{\ell+1, \ell+k}^{(N)} \right) \quad (k = 2, \dots, n; \ell = N, N-1, \dots, 0), \quad (72)$$

$$R_{\ell, \ell+k} = \lim_{N \rightarrow \infty} R_{\ell, \ell+k}^{(N)} \quad (k = 1, \dots, n; \ell = 0, 1, 2, \dots). \quad (73)$$

From a numerical point of view, *BA* requires less computational effort than *FA*. The only disadvantage of *BA* is that for each convergent, the whole algorithm has to be repeated which can be time-consuming if the algorithm converges slowly. Since we are not interested in a particular $R_{i,i+k}$ but in the family of matrices $R_{i,i+k}$, we take *BA* for the preferable algorithm in all situations.

It is easy to see that the recurrence relations (71)–(73) are equivalent to (12) and (13) with $R_{\ell,\ell+k} = R_\ell \cdot R_{\ell+1} \cdots R_{\ell+k-1}$. Hence, the derivation based on continued fractions leads to the same results as the derivation based on taboo probabilities.

Observe that the derivation based on taboo probabilities still works if $\det(P_{i,i+1}) \neq 0$ for some i , whereas the derivation based on matrix continued fractions fails. Nevertheless, the derivation based on matrix continued fractions enables us to detect that (65) is a decoupled version of the difference equation (40) which is important with respect to numerical applications. In [31], it was shown that for $n = 1$ and $d = 1$, the difference equation (37) is subject to numerical instability, because the invariant measure \mathbf{x} is asymptotically dominated by other solutions of the difference equation (see [32]). In addition, it was shown that the continued fraction algorithms (65) and (71) result in a decoupled recursion for the computation of \mathbf{x} in which the dominant solutions of (37) do not appear. But however, we cannot assure that in general, the decoupled recursion (65) is kept free from all sources of numerical instability. To take this issue forward, we need tailored Pringsheim- and Pincherle-type convergence criteria for matrix continued fractions, compare with [29].

Remark 4. The existence of the limits (68) has been shown in Theorem 1 by probabilistic arguments. To prove (68) in a purely analytic manner, further research is required. The reason is the following: Obviously, the convergence of the matrix-driven Jacobi-Perron algorithm is implied by the diagonal dominance of the transition probability matrix P . In general, there exists a great variety of convergence criteria for ordinary continued fractions and their generalizations which appeal to the diagonal dominance of their coefficients (the so-called Pringsheim criteria). But in the case of the matrix-driven Jacobi-Perron algorithm, these results are only available in terms of norm-based criteria, which are too restrictive to imply convergence in the present case. In so far, it may be worth to study the relationship between Markov chains and approximation theory in more detail in order to discover alternate techniques for detecting convergence of the matrix-driven Jacobi-Perron algorithm. Notice that the approximants of the invariant measure \mathbf{x} of (1) are entirely determined by the finite difference schemes (65)–(67). In so far, the analytic proof of the existence of the limits (68) is mainly of theoretical interest.

6. Numerical Example

Consider a queueing system consisting of a single server which is subject to breakdowns. The up time of the system

is exponentially distributed with parameter $\alpha > 0$. The down time is assumed to be Erlang distributed with shape parameter $K \in \mathbb{N}$ and rate parameter $K\beta$ for some $\beta > 0$ (implying that the mean down time is $1/\beta$). Customers arrive according to a Poisson stream with rate $\lambda > 0$. By $\kappa_i > 0$, we denote the probability that an arriving customer joins the queue when i customers are present and the server is active. Analogously, $\theta_i \geq 0$ shall denote the probability that an arriving customer joins the queue when i customers are present and the server is down. Notice that it could make sense to assume $\theta_i = 0$, that is, customers will not join the queue when the system is interrupted. When less than n customers are present, no service is provided. Otherwise, n customers are served at a time. The service time of each batch is assumed to be exponentially distributed with parameter $\mu > 0$. In addition, we assume that the arrival process, the service process, and the breakdown mechanism are stochastically independent. The model under consideration is an extension of the usual $M/M^{(n,n)}/1$ queue, which is described in more detail in [33].

Remember that an Erlang distributed random variable may be interpreted as a phase-type distributed random variable [1] with distribution function $F(x) = 1 - \alpha \exp(-Tx)\mathbf{1}$ for $x \geq 0$, where

$$T = \begin{pmatrix} -K\beta & K\beta & & & \\ & -K\beta & K\beta & & \\ & & \ddots & \ddots & \\ & & & -K\beta & K\beta \\ & & & & -K\beta \end{pmatrix} \text{ and } \alpha = (1, 0, \dots, 0). \quad (74)$$

Erlang distributions are frequently used when modelling random variables with a small variance or even approximating deterministic behaviour. This is due to the fact that the variance of the Erlang distribution with parameters K and $K\beta$ is $1/K\beta^2$, and this is indeed the smallest variance which can occur for a phase-type distribution with order K and expectation $1/\beta$ (see [34] for more details).

Due to the lack-of-memory property of the exponential distribution and the independence assumptions, the system under consideration is characterized by a two-dimensional continuous-time Markov chain $Y = (Y_t)_{t \geq 0}$ with $Y_t = (N_t, U_t)$ where

- (i) N_t is the number of customers in the system at time t
- (ii) $U_t = j \in \{1, \dots, K\}$ indicates that the server is down and the down time is in phase j at time t
- (iii) $U_t = K + 1$ indicates that the server is up at time t .

Hence, the underlying state space is $E = \{0, 1, 2, \dots\} \times \{1, 2, \dots, K + 1\}$, and we obtain a generator matrix of structure (20) with

$$\begin{aligned}
Q_{i,i+1} &= \begin{pmatrix} \lambda\theta_i & 0 & \cdots & 0 & 0 \\ 0 & \lambda\theta_i & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda\theta_i & 0 \\ 0 & \cdots & \cdots & 0 & \lambda\kappa_i \end{pmatrix} \quad (i=0, 1, 2, \dots), \\
Q_{i,i-n} &= \begin{pmatrix} 0 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & \mu & \end{pmatrix} \quad (i=n, n+1, \dots), \\
Q_{i,i} &= \begin{pmatrix} -\lambda\theta_i - K\beta & K\beta & & & \\ & -\lambda\theta_i - K\beta & K\beta & & \\ & & \ddots & \ddots & \\ & & & -\lambda\theta_i - K\beta & K\beta \\ \alpha & 0 & \cdots & 0 & -\lambda\kappa_i - \alpha \end{pmatrix} \\
&\quad (i=0, 1, \dots, n-1), \\
Q_{i,i} &= \begin{pmatrix} -\lambda\theta_i - K\beta & K\beta & & & \\ & -\lambda\theta_i - K\beta & K\beta & & \\ & & \ddots & \ddots & \\ & & & -\lambda\theta_i - K\beta & K\beta \\ \alpha & 0 & \cdots & 0 & -\lambda\kappa_i - \alpha - \mu \end{pmatrix} \\
&\quad (i=n, n+1, \dots).
\end{aligned} \tag{75}$$

A sufficient condition for Y to be regular and ergodic is

$$\limsup_{i+j \rightarrow \infty} \lambda \cdot \max \{\theta_i, \kappa_i\} < n\mu \cdot \frac{\beta}{\alpha + \beta}. \tag{76}$$

This statement can be deduced from Tweedie's criterion ([35], Theorem 2.3). The condition is to find a nonnegative sequence $\xi_{(i,j)}$ and an integer $N > 0$ satisfying $\xi_{(i,j)} \rightarrow \infty$ for $i+j \rightarrow \infty$ and

$$\begin{aligned}
\lambda\theta_i [\xi_{(i,j)} - \xi_{(i+1,j)}] + K\beta [\xi_{(i,j)} - \xi_{(i,j+1)}] &\geq \varepsilon, \quad j=1, \dots, K, \\
\lambda\kappa_i [\xi_{(i,K+1)} - \xi_{(i+1,K+1)}] + \alpha [\xi_{(i,K+1)} - \xi_{(i,0)}] \\
+ \mu [\xi_{(i,K+1)} - \xi_{(i-n,K+1)}] &\geq \varepsilon,
\end{aligned} \tag{77}$$

$i \geq n, i+j \geq N$, and some $\varepsilon > 0$. A simple check shows that

$$\xi_{(i,j)} = i + \frac{K+1-j}{K} \cdot \frac{n\mu}{\alpha + \beta} \quad (i=0, 1, 2, \dots; j=1, 2, \dots, K+1), \tag{78}$$

satisfies this set of inequalities as long as (76) holds.

Next, we focus on the calculation of the following performance values:

- (i) The expected stationary number $\mathbb{E}[N_\infty]$ of customers in the system is given by

$$\mathbb{E}[N_\infty] = \sum_{i=0}^{\infty} \pi_i \cdot \begin{pmatrix} i \\ i \\ \vdots \\ i \\ i \end{pmatrix}, \tag{79}$$

where $\pi_i = (\pi_{(i,1)}, \dots, \pi_{(i,K+1)})$

- (ii) The corresponding variance is calculated according to $\text{VAR}[N_\infty] = \mathbb{E}[N_\infty^2] - \mathbb{E}[N_\infty]^2$ where

$$\mathbb{E}[N_\infty^2] = \sum_{i=0}^{\infty} \pi_i \cdot \begin{pmatrix} i^2 \\ i^2 \\ \vdots \\ i^2 \\ i^2 \end{pmatrix} \tag{80}$$

- (iii) The stationary probability that no service is provided despite $\geq n$ customers waiting is given by

$$\tilde{p} := \mathbb{P}(U_\infty \leq K \mid N_\infty \geq n) = \frac{\mathbb{P}(N_\infty \geq n, U_\infty \leq K)}{\mathbb{P}(N_\infty \geq n)}, \tag{81}$$

where

$$\begin{aligned}
\mathbb{P}(N_\infty \geq n, U_\infty \leq K) &= \sum_{i=n}^{\infty} \sum_{j=1}^K \pi_{(i,j)} = \sum_{i=n}^{\infty} \pi_i \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \\
\mathbb{P}(N_\infty \geq n) &= \sum_{i=n}^{\infty} \sum_{j=1}^{K+1} \pi_{(i,j)} = \sum_{i=n}^{\infty} \pi_i \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}
\end{aligned} \tag{82}$$

TABLE 1: $\mathbb{E}[N_\infty]$ for $\lambda = 1.0$ (a) and $\lambda = 4.0$ (b), respectively.

(a)				
$\alpha \backslash \beta$	2	1	0.5	0.1
0.5	1.825	1.855	1.925	2.537
0.1	1.806	1.813	1.832	2.096
0.05	1.803	1.807	1.817	1.985
0.01	1.801	1.802	1.804	1.848

(b)				
$\alpha \backslash \beta$	2	1	0.5	0.1
0.5	4.413	4.531	4.787	6.912
0.1	4.332	4.358	4.423	5.192
0.05	4.322	4.335	4.369	4.844
0.01	4.314	4.316	4.323	4.445

TABLE 2: $\text{VAR}[N_\infty]$ for $\lambda = 1.0$ (a) and $\lambda = 4.0$ (b), respectively.

(a)				
$\alpha \backslash \beta$	2	1	0.5	0.1
0.5	1.853	1.892	1.977	2.798
0.1	1.826	1.834	1.854	2.132
0.05	1.822	1.826	1.837	2.014
0.01	1.819	1.820	1.822	1.871

(b)				
$\alpha \backslash \beta$	2	1	0.5	0.1
0.5	6.969	7.155	7.544	11.636
0.1	6.818	6.854	6.927	7.941
0.05	6.800	6.817	6.853	7.462
0.01	6.785	6.789	6.796	6.959

(iv) The stationary probability of the system being down is

$$\mathbb{P}(U_\infty \leq K) = \sum_{i=0}^{\infty} \sum_{j=1}^K \pi_{(i,j)} = \sum_{i=0}^{\infty} \pi_i \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix}. \quad (83)$$

The latter probability can be computed exactly as $1/\beta/(\frac{1}{\alpha} + 1/\beta) = \alpha/(\alpha + \beta)$. Therefore, the numerical computation of this probability only helps us to check the implementation.

For calculating all these values, we apply Algorithm 2 with $L = 5$ and

$$f(i) = \begin{pmatrix} i & i^2 & 1_{\{n,n+1,\infty\}}(i) & 1_{\{n,n+1,\infty\}}(i) & 1 & 1 \\ i & i^2 & 1_{\{n,n+1,\infty\}}(i) & 1_{\{n,n+1,\infty\}}(i) & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ i & i^2 & 1_{\{n,n+1,\infty\}}(i) & 1_{\{n,n+1,\infty\}}(i) & 1 & 1 \\ i & i^2 & 0 & 1_{\{n,n+1,\infty\}}(i) & 0 & 1 \end{pmatrix}, \quad (84)$$

where $1_A(\mathbf{x}) = 1$ for $\mathbf{x} \in A$ and $1_A(\mathbf{x}) = 0$ for $\mathbf{x} \notin A$. After applying the algorithm, the variance can be deduced from the first two entries of the resulting vector. Similarly, the conditional probability $\tilde{p} = \mathbb{P}(U_\infty \leq K \mid N_\infty \geq n)$ can be obtained from the third and the forth entry.

In Tables 1, 2, and 3, we present some numerical results for $n = 3$, $K = 10$, $\mu = 1.0$, $\lambda = 1.0$ or $\lambda = 4.0$, $\kappa_i = 1/(\lfloor i/n \rfloor + 1)$, and $\theta_i = \kappa_i/10$ for $i = 0, 1, 2, \dots$, and various values of α and β .

TABLE 3: \tilde{p} for $\lambda = 1.0$ (a) and $\lambda = 4.0$ (b), respectively.

(a)				
$\alpha \backslash \beta$	2	1	0.5	0.1
0.5	0.204	0.344	0.523	0.868
0.1	0.0488	0.0952	0.181	0.592
0.05	0.0250	0.0500	0.0994	0.424
0.01	0.00511	0.0104	0.0216	0.130

(b)				
$\alpha \backslash \beta$	2	1	0.5	0.1
0.5	0.202	0.340	0.512	0.846
0.1	0.0483	0.0935	0.175	0.539
0.05	0.0248	0.0490	0.0958	0.372
0.01	0.00506	0.0102	0.0208	0.107

TABLE 4: $\mathbb{E}[N_\infty]$, $\text{VAR}[N_\infty]$, and \tilde{p} as a function of the coefficient of variation of the down time (i.e., $\text{CV} = 1/\sqrt{K}$).

K	1	3	5	10	20
$\mathbb{E}[N_\infty]$	2.244	2.136	2.113	2.096	2.087
$\text{VAR}[N_\infty]$	2.431	2.200	2.160	2.132	2.118
p_u	0.631	0.604	0.598	0.592	0.589

In Table 4, we fix $\lambda = \mu = 1.0$ and $\alpha = \beta = 0.1$ and vary the coefficient of variation of the down time (i.e., $\text{CV} = 1/\sqrt{K}$) from 1 to $1/\sqrt{20}$.

The usual $M/G/1$ queue with service interruptions has been investigated in great detail by Gaver [36] and Federgruen and Green [37]. Their explicit formulas tell us that the mean number of customers in the system and the variance of the number of customers in the system are strongly affected by the arrival rate (i.e., λ), the ratio of the mean repair time and the mean service time (i.e., μ/β), and the

availability of the system (i.e., $\mathbb{P}(U_\infty = K + 1) = \beta/(\alpha + \beta)$). Obviously, the same is true for the state dependent bulk service queue with unreliable server (see Tables 1 and 2).

The conditional probability \tilde{p} may be interpreted as a certain measure for the non-value added time of the server. The data in Table 3 suggest that for $\lambda \rightarrow \infty$, convergence is to $\alpha/(\alpha + \beta) = \mathbb{P}(U_\infty \leq K)$. This can be deduced from the law of total probability which states

$$\begin{aligned} \frac{\alpha}{\alpha + \beta} = \mathbb{P}(U_\infty \leq K) &= \tilde{p} \cdot \mathbb{P}(N_\infty \geq n) \\ &+ \mathbb{P}(U_\infty \leq K \mid N_\infty < n) \cdot (1 - \mathbb{P}(N_\infty \geq n)). \end{aligned} \quad (85)$$

Since $\mathbb{P}(N_\infty \geq n) \rightarrow 1$ for $\lambda \rightarrow \infty$, we conclude from (85) that $\tilde{p} \rightarrow \mathbb{P}(U_\infty \leq K) = \alpha/(\alpha + \beta)$ as $\lambda \rightarrow \infty$. So it is not surprising that α and β are the main influencing factors, whereas the arrival rate λ has only little effect on \tilde{p} which may be explained as follows: Due to its state dependence, the arrival stream in our example is massively dampened when the number of customers grows. In other words, in spite of an excessive arrival rate (i.e., $\lambda > \mu$), the system always reaches a comparatively moderate and balanced state.

Finally, Table 4 reveals that the coefficient of variation of the down time has a relatively small influence on the main characteristics of the system.

It should be mentioned that our model allows the specification of more complex interarrival, service, and up and down times by fitting phase-type distributions to empirical data or by approximating general distributions by phase-type distributions. Since it is only our intention to demonstrate the practical suitability of our algorithm, we do not deepen the discussion of the numerical results.

Obviously, a lot of other queueing situations can be taken into account within our model. In [17], a tandem queue was analysed in which

- (i) customers arrive at the first station according to a Markovian arrival process
- (ii) customers served at the first station are transferred to the second station
- (iii) customers served there leave the system
- (iv) there are c_1 and c_2 servers at both stations, respectively
- (v) at both stations, the service time is phase-type distributed.

As pointed out in [17], the state space of the underlying queueing process consists of the number of waiting customers (at the first station and at the second station), the number of servers serving in the various phases of the phase-type distribution (at the first station and at the second station), and the environmental state for the Markovian arrival process. But by appropriately renumbering the states, it is seen that the process may be transformed into a quasi-

birth-death process having a generator matrix of the form (20) with $n = 1$.

Next, consider a batch service at both stations with a maximum batch size n at the second station. Then departures at the second station decrease the total number of customers in the system by n . Hence, a structure of type (20) with non-constant block-dimensions comes into play. Since the state space remains finite, algorithms 1 and 2 are exact up to numerical errors. Similarly, many other queueing models may be solved in this manner.

7. Conclusion and Further Research

In this paper, we have suggested methods for computing invariant measures and stationary expectations for Markov chains with a band-block-structured transition probability matrix or generator matrix, respectively. We have presented a probabilistic derivation as well as a derivation based on generalized matrix-valued continued fractions. Although the latter derivation only works under some restrictive assumptions, we have gained some additional numerical and algebraic insights. From a practical point of view, we want to emphasize the benefits of Algorithm 2. This method generalizes the established algorithm for computing stationary expectations for quasi-birth-death processes [7], and in situations where we are only interested in stationary expectations (what should be the case in most practical applications), it is extremely efficient with respect to memory requirement.

As pointed out in the remarks concerning the Algorithms 1 and 2, all considerations but the memory-efficient computation of stationary expectations can be easily extended to lower-block Hessenberg matrices. Unfortunately, for general lower-block Hessenberg matrices (and large truncation level N), the memory requirement will be crucial. Hence, further research could concern memory-efficient use of the lower-block Hessenberg transition structure (skip-free to the right). Other structures being of interest in practical applications are upper-block Hessenberg matrices or general block-band matrices with upper bandwidth > 1 . Another interesting point concerns the choice of the truncation level N and the derivation of precise error bounds.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We acknowledge support by Open Access Publishing Fund of Clausthal University of Technology.

References

- [1] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, The John Hopkins University Press, Baltimore, 1981.

- [2] S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes*, Academic Press, New York, 1981.
- [3] L. Bright and P. G. Taylor, "Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes," *Communications in Statistics. Stochastic Models*, vol. 11, no. 3, pp. 497–525, 1995.
- [4] T. Phung-Duc, H. Masuyama, S. Kasahara, and Y. Takahashi, "A simple algorithm for the rate matrices of level-dependent qbd processes," in *Proceedings of the 5th International Conference on Queueing Theory and Network Applications - QTNA '10*, pp. 46–52, Beijing, China, 2010.
- [5] T. Hanschke, "A matrix continued fraction algorithm for the multiserver repeated order queue," *Mathematical and Computer Modelling*, vol. 30, no. 3–4, pp. 159–170, 1999.
- [6] H. Baumann and W. Sandmann, "Numerical solution of level dependent quasi-birth-and-death processes," *Procedia Computer Science*, vol. 1, no. 1, pp. 1561–1569, 2010.
- [7] H. Baumann and W. Sandmann, "Computing stationary expectations in level-dependent QBD processes," *Journal of Applied Probability*, vol. 50, no. 1, pp. 151–165, 2013.
- [8] H. Masuyama, "Limit formulas for the normalized fundamental matrix of the northwestcorner truncation of Markov chains: matrix-infinite-product-form solutions of blockhesenberg Markov chains. Technical report," Kyoto University, 2018.
- [9] E. Seneta, "Finite approximations to infinite non-negative matrices," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 63, no. 4, pp. 983–992, 1967.
- [10] E. Seneta, "Finite approximations to infinite non-negative matrices, II: refinements and applications," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 64, no. 2, pp. 465–470, 1968.
- [11] D. G. Kendall and G. E. H. Reuter, "The calculation of the ergodic projection for Markov chains and processes with a countable infinity of states," *Acta Mathematica*, vol. 97, pp. 103–144, 1957.
- [12] G. E. H. Reuter, "Denumerable Markov processes and the associated contraction semigroups on l_1 ," *Acta Mathematica*, vol. 97, pp. 1–46, 1957.
- [13] T. Dayar, W. Sandmann, D. Spieler, and V. Wolf, "Infinite level-dependent QBD processes and matrix-analytic solutions for stochastic chemical kinetics," *Advances in Applied Probability*, vol. 43, no. 4, pp. 1005–1026, 2011.
- [14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The John Hopkins University Press, 1996.
- [15] A. Ost, *A Performance of Communication Systems: A Model-Based Approach with Matrix Geometric Methods*, Springer Science & Business Media, 2013.
- [16] H. Baumann and W. Sandmann, "Bounded truncation error for long-run averages in infinite Markov chains," *Journal of Applied Probability*, vol. 52, no. 3, pp. 609–621, 2015.
- [17] H. Baumann and W. Sandmann, "Multi-server tandem queue with Markovian arrival process, phase-type service times, and finite buffers," *European Journal of Operational Research*, vol. 256, no. 1, pp. 187–195, 2017.
- [18] M. G. de Bruin, "Generalized C-fractions and a multi-dimensional Padé table. Doctoral Thesis," University of Amsterdam, 1974.
- [19] P. van der Cruyssen, "Discussion of algorithms for the computation of generalized continued fractions," *Journal of Computational and Applied Mathematics*, vol. 8, no. 3, pp. 179–2186, 1982.
- [20] O. Perron, *Die Lehre von den Kettenbrüchen*, Teubner, Stuttgart, 1954.
- [21] C. G. J. Jacobi, "Allgemeine Theorie der kettenbruchähnlichen Algorithmen, in welchen jede Zahl aus drei vorhergehenden gebildet wird," *Journal für die Reine und Angewandte Mathematik*, vol. 1868, no. 69, pp. 29–64, 1868.
- [22] O. Perron, "Grundlagen für eine Theorie des Jacobischen Kettenbruchalgorithmus," *Mathematische Annalen*, vol. 64, no. 1, pp. 1–76, 1907.
- [23] B. P. Bachmann, "Zur Theorie von Jacobis Kettenbruch-Algorithmen," *Journal für die Reine und Angewandte Mathematik (Crelles Journal)*, vol. 1873, no. 75, pp. 25–34, 1873.
- [24] L. Bernstein, *The Jacobi-Perron Algorithm: Its Theory and Application*, vol. 207 of Lecture Notes in Mathematics, Springer, 1971.
- [25] F. Schweiger, *The Metrical Theory of Jacobi-Perron Algorithm*, vol. 334 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1973.
- [26] T. Hanschke, "Markov chains and generalized continued fractions," *Journal of Applied Probability*, vol. 29, no. 4, pp. 838–849, 1992.
- [27] W. Groebner, *Matrizenrechnung*, Bibliographisches Institut, Mannheim, 1966.
- [28] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, Academic Press London, Second edition, 1985.
- [29] P. Levrie and A. Bultheel, "Matrix continued fractions related to first-order linear recurrence systems," *Electronic Transactions on Numerical Analysis*, vol. 4, pp. 46–63, 1996.
- [30] O. Perron, "Über die Konvergenz der Jacobi-Kettenalgorithmen mit komplexen Elementen," *Sitzungsberichte der mathematisch-physikalischen Klasse der Königlich Bayerischen Akademie der Wissenschaften zu München*, vol. 37, pp. 401–482, 1907.
- [31] H. Baumann and T. Hanschke, "Inherent numerical instability in computing invariant measures of Markov chains," *Applied Mathematics*, vol. 8, no. 9, pp. 1367–1385, 2017.
- [32] W. Gautschi, "Zur Numerik rekurrenter Relationen," *Computing*, vol. 9, no. 2, pp. 107–126, 1972.
- [33] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, John Wiley & Sons, Inc., 1985.
- [34] D. Aldous and L. Shepp, "The least variable phase type distribution is Erlang," *Communications in Statistics. Stochastic Models*, vol. 3, no. 3, pp. 467–473, 1987.
- [35] R. L. Tweedie, "Sufficient conditions for regularity, recurrence and ergodicity of Markov processes," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 78, no. 1, pp. 125–136, 1975.
- [36] D. P. Gaver Jr., "A waiting line with interrupted service, including priorities," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 24, no. 1, pp. 73–90, 1962.
- [37] A. Federgruen and L. Green, "Queueing systems with service interruptions," *Operations Research*, vol. 34, no. 5, pp. 752–768, 1986.